15/6/23 - AN.

Reg. No. : | | | | | | | | | | | | |

Question Paper Code : 50040

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2023.

Sixth Semester

Artificial Intelligence and Data Science

AD 8006 – ENGINEERING PREDICTIVE ANALYTICS

(Regulations 2017)

Time : Three hours Maximum : 100 marks

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. Differentiate between predictive analytics and data analytics.

2. Define Data Audit.

3. Explain the two goals needed in preparing the data for better performance and scalability.

4. List the two interesting measures of an association rule.

5. Write the challenges in clustering.

6. Classify the data preparation issues with descriptive model.

7. Point out the responsibilities of K-Nearest Neighbors algorithm.

8. Assess the ways to deal neural network model.

9. Why Text Mining Is Hard? Give reasons.

10. Give the structure of regular expressions.

PART B — (5 × 13 = 65 marks)

11. (a) Define data visualization. Explain the different data dimensions in visualizing with an example.

Or

(b) What do you mean by statistical significance? Discuss the importance of predictive analytics. Give example.

12. (a) Describe the working principle of building classification rules from association rules with a real time scenario.

Or

(b) (i) Illustrate the process of preparing data for association rules in detail. (7)

(ii) Explain the terms feature creation and parameter setting in data preparation process. (6)

13. (a) Define the term principal component analysis. Develop the design cycle of principal component analysis and explain components involved with an example.

Or

(b) Discuss in detail about the various standard cluster model interpretations.

14. (a) (i) Can linear regression be used for classification. Justify. (7)

(ii) Express the Naive Bayes modelling with examples (6)

Or

(b) The following table consists of training data from an employee database. The data have been generalized. Let status be the class label attribute. Construct Decision tree from the given data.

Department	Age	Salary	Count	Status
Sales	31...35	46k...50k	30	Senior
Sales	26...30	26k...30k	40	Junior
Sales	31...35	31k...35k	40	Junior
Systems	21...25	46k...50k	20	Junior
Systems	31...35	66k...70k	5	Senior
Systems	26...30	46k...50k	3	Junior
Systems	41...35	66k...70k	3	Senior
Marketing	36...40	46k...50k	10	Senior
Marketing	31...35	41k...45k	4	Junior
Secretary	46...50	36k...40k	4	Senior
Secretary	26...30	26k...30k	6	Junior

15. (a) Discuss the Predictive Modeling Approach to Text Mining. In what way Structured vs. Unstructured Data helps the mining process.

Or

(b) List out the Data Preparation Steps for text mining. Elaborate on Modeling with Text Mining Features.

PART C — (1 × 15 = 15 marks)

16. (a) Consider the training dataset given in the following table. Use Weighted k-NN and determine the class. Test instance (7.6, 60, 8) and K = 3.

S.No.	CGPA	Assessment	Project Submitted	Result
1	9.2	85	8	Pass
2	8	80	7	Pass
3	8.5	81	8	Pass
4	6	45	5	Fail
5	6.5	50	4	Fail
6	8.2	72	7	Pass
7	5.8	38	5	Fail
8	8.9	91	9	Pass

Or

(b) Consider a home finance loan to predict the housing loan payment. Design a general hierarchical structure and analyze the factors using rule discovery techniques to accurately predict the number of loan payments in a given quarter/year. A loan is availed for a period of 20 to 25 years, but the average life span of the loan exists for only 7 to 10 years due to payment. Make necessary assumptions: Maintenance record of the customer details and details of the prevailing interest rates, borrower characteristics, account dare, fine tunc loan prepayment such as interest rates and fees to maximize the profits of the company. Elaborately discuss the association rule mining issues. Also Examine on the multi-level association rules and find if you could relate relation on from the above application.

———————